Volume 8 Issue 4 (2024) Pages 687-698
Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini
ISSN: 2549-8959 (Online) 2356-1327 (Print)

Pengujian Instrumen Akreditasi PAUD dengan Model Rasch

Petrus Redy Partus Jaya✉, Beata Palmin[2], Theresia Alviani Sum[3]
Pendidikan Guru Pendidikan Anak Usia Dini, Universitas Katolik Indonesia Santu Paulus Ruteng, Indonesia[(1,2,3)]
DOI: 10.31004/obsesi.v8i4.5953

Abstrak

Kualitas Pendidikan Anak Usia Dini (PAUD) sangat dipengaruhi oleh instrumen akreditasi yang valid dan reliabel, yang dapat memberikan penilaian objektif terhadap stimulasi kognitif yang diberikan oleh guru. Penelitian ini bertujuan untuk mengevaluasi dan mengkritisi validitas serta reliabilitas instrumen visitasi akreditasi PAUD menggunakan Model Rasch 1 Parameter Logistic (1PL). Penelitian ini menggunakan pendekatan kuantitatif dengan melibatkan 124 lembaga PAUD dan 47 asesor pada tahun 2023. Data dikumpulkan melalui penilaian performa guru dalam menstimulasi aspek kognitif anak usia dini. Analisis data dilakukan menggunakan Model Rasch yang mengevaluasi kesesuaian item dengan model serta variasi tingkat kesulitan item. Hasil penelitian menunjukkan bahwa sebagian besar item dalam instrumen sesuai dengan model Rasch dan memiliki variasi tingkat kesulitan yang memadai. Beberapa item diidentifikasi memerlukan revisi untuk meningkatkan kualitas instrumen. Temuan ini menegaskan validitas dan reliabilitas instrumen, serta memberikan rekomendasi perbaikan untuk akreditasi PAUD di Indonesia. Implikasi penelitian mencakup peningkatan konsistensi dan akurasi penilaian melalui penyempurnaan instrumen dan pelatihan asesor.
Kata Kunci: *model rasch; akreditasi paud; stimulasi kognitif; validitas; reliabilitas.*

Abstract

The quality of Early Childhood Education (ECE) is greatly influenced by the validity and reliability of accreditation instruments, which provide objective assessments of the cognitive stimulation offered by teachers. This study aims to evaluate and critique the validity and reliability of the accreditation visitation instrument for ECE using the Rasch Model 1 Parameter Logistic (1PL). A quantitative approach was employed, involving 124 ECE institutions and 47 assessors in 2023. Data were collected through assessments of teachers' performance in stimulating the cognitive aspects of young children. Data analysis was conducted using the Rasch Model to evaluate item fit and the range of item difficulty. The findings indicate that most items in the instrument align well with the Rasch model and exhibit an adequate range of difficulty levels. However, several items were identified as needing revision to enhance the overall quality of the instrument. These findings affirm the validity and reliability of the instrument and provide recommendations for improving ECE accreditation in Indonesia. The study's implications include enhancing the consistency and accuracy of assessments through instrument refinement and additional training for assessors.
Keywords: *Rasch Model; ECE accreditation; cognitive stimulation; validity; reliability.*

✉ Corresponding author :
Email Address : petrusredypartusjaya@gmail.com (alamat koresponden)
Received 7 June 2024, Accepted 13 August 2024, Published 13 August 2024

DOI: 10.31004/obsesi.v84.5953

Pendahuluan

Kualitas PAUD di Indonesia masih memerlukan peningkatan. Berdasarkan hasil akreditasi satuan PAUD tahun 2019, dari 34.762 satuan PAUD yang dinilai, hanya 8,55% yang terakreditasi A, sementara mayoritas berada di tingkat akreditasi B dan C (Kemendikbud, 2022). Ini menunjukkan bahwa masih banyak satuan PAUD yang belum mencapai tingkat mutu yang optimal. Selain itu, analisis *Early Childhood Development Index* (ECDI) menunjukkan bahwa meskipun anak usia dini di Indonesia memiliki kesehatan fisik yang baik, kemampuan literasi dan numerasi mereka masih perlu ditingkatkan (Kemendikbud, 2022). Untuk mengatasi tantangan ini, pemerintah telah melakukan berbagai upaya, termasuk pengembangan instrumen penilaian yang lebih komprehensif dan integrasi data untuk penilaian kinerja satuan pendidikan. Sejak tahun 2021 melalui keputusan Menteri Pendidikan dan Kebudayaan Nomor 71/p/2021, instrumen penilaian akreditasi satuan PAUD terintegrasi dalam aplikasi Sistem Penilaian Akreditasi (SISPENA 3.1). Penggunaan instrumen ini diharapkan dapat mengukur kualitas satuan PAUD secara valid dan reliabel. Meskipun demikian, dalam pengukuran mutu PAUD perlu diperhatikan bahwa instrumen penilaian yang digunakan mesti memberikan hasil yang objektif, valid, dan reliabel. Untuk itu, perlu dilakukan pengujian terhadap kualitas instrumen yang digunakan.

Dalam instrumen SISPENA PAUD 3.1, skala guttman digunakan untuk mengidentifikasi pemenuhan standar mutu yang telah ditetapkan. Dalam konteks ini, Skala Guttman muncul sebagai salah satu metode pengukuran yang digunakan untuk menilai kualitas satuan PAUD (Frey, 2018; Roßbach, 1990). Skala Guttman, yang juga dikenal sebagai skala dikotomis, diharapkan dapat menghasilkan data yang lebih presisi dan akurat yang memungkinkan analisis mendalam, serta memberikan informasi detail mengenai kekuatan dan kelemahan institusi PAUD. Namun, sebagaimana instrumen penilaian lainnya, penggunaan Skala Guttman tidak lepas dari potensi bias (Clark & Kriedt, 1948; Pillet et al., 2018). Kesalahan bias dapat memengaruhi validitas dan reliabilitas hasil penilaian (Lash et al., 2021; McGrath et al., 2010), yang pada akhirnya memengaruhi keputusan yang diambil berdasarkan hasil tersebut.

Fakta ini mendorong para peneliti untuk melakukan penelitian evaluasi terkait potensi kesalahan bias dalam instrumen penilaian akreditasi PAUD menggunakan Skala Guttman. Mengingat urgensi dan relevansinya, penelitian ini dirancang untuk menyelidiki kesalahan bias potensial dalam instrumen penilaian akreditasi PAUD yang menggunakan Skala Guttman. Hasil penelitian ini diharapkan dapat memberikan rekomendasi untuk perbaikan instrumen penilaian tersebut, sehingga proses akreditasi PAUD dapat dilakukan dengan lebih akurat dan objektif.

Penelitian ini berada di garis depan dalam mengevaluasi instrumen visitasi akreditasi PAUD dengan menggunakan Model Rasch 1 Parameter Logistic (1PL)(A Aziz et al., 2019; Umobong, 2017). Model Rasch dikenal luas dalam analisis item (de Morton & Keating, 2008; Karlin & Karlin, 2018; Sirodj et al., 2020) karena kemampuannya untuk memberikan informasi mendalam mengenai kesesuaian item dan responden, serta mengidentifikasi item yang memerlukan perbaikan. Dengan aplikasi Model Rasch, penelitian ini tidak hanya berupaya menilai validitas dan reliabilitas instrumen akreditasi, tetapi juga memberikan wawasan penting untuk perbaikan instrumen tersebut (Krishnan & Idris, 2018; Mohamad et al., 2015). Pendekatan ini memberikan analisis yang lebih akurat dan objektif (Boone & Rogan, 2005; Yan, 2010) dibandingkan metode lainnya seperti Skala Guttman, yang lebih rentan terhadap bias. Dengan menggunakan Model Rasch, penelitian ini berupaya memastikan bahwa setiap item dalam instrumen mampu secara efektif mengukur performa guru (Wilkerson & Lang, 2006) dalam memberikan stimulasi kognitif kepada anak-anak (Bui et al., 2020)("Appl. Rasch Meas. Learn. Environ. Res.," 2011) .

Fokus utama evaluasi dalam penelitian ini adalah pada aspek stimulasi kognitif anak usia dini, yang merupakan komponen penting dalam perkembangan anak. Stimulasi kognitif mencakup berbagai kegiatan yang mendorong anak untuk berpikir kritis, memecahkan

DOI: 10.31004/obsesi.v84.5953

masalah, dan memahami konsep-konsep dasar yang esensial bagi perkembangan kognitif mereka. Penelitian ini mengevaluasi bagaimana guru menstimulasi anak untuk mengidentifikasi masalah sesuai dengan tingkat usia mereka (item 1), mencari solusi atas masalah tersebut (item 2), serta memberikan dukungan (scaffolding) yang tepat (item 3). Selain itu, penelitian ini juga menilai bagaimana guru memberikan penghargaan atas keberhasilan anak (item 4), serta menstimulasi anak dalam memahami persamaan (item 5), perbedaan (item 6), dan konsep menghubungkan (item 7). Lebih lanjut, evaluasi dilakukan terhadap cara guru membantu anak memahami konsep menghubungkan (item 8), mengklasifikasikan objek berdasarkan bentuk, ukuran, dan warna (item 9), memahami konsep sebab-akibat (item 10), mengenal (item 11) serta menggunakan konsep bilangan dan huruf (item 12), hingga mengekspresikan ide dan imajinasi mereka melalui berbagai bentuk karya (item 13). Dengan fokus pada aspek-aspek ini, penelitian ini bertujuan untuk memastikan bahwa instrumen visitasi akreditasi PAUD dapat mengukur stimulasi kognitif secara efektif dan akurat. Hasil evaluasi ini diharapkan dapat digunakan sebagai alat yang valid dan reliabel dalam penilaian akreditasi PAUD, sehingga dapat memberikan kontribusi nyata dalam peningkatan kualitas pendidikan anak usia dini di Indonesia.

Metodologi

Pendekatan penelitian yang digunakan dalam studi ini adalah pendekatan kuantitatif. Data dikumpulkan dari 124 lembaga Pendidikan Anak Usia Dini (PAUD) yang dinilai oleh 47 asesor dalam penilaian visitasi pada tahun 2023. Instrumen penilaian visitasi yang digunakan terdiri dari berbagai item yang mengukur aspek performa guru dalam menstimulasi aspek kognitif anak usia dini. Untuk memastikan validitas dan reliabilitas instrumen tersebut, analisis data dilakukan menggunakan Model Rasch 1 Parameter Logistic (1 PL). Model Rasch 1 PL, juga dikenal sebagai Model Rasch sederhana, hanya memperhitungkan tingkat kesulitan item sebagai parameter, dengan asumsi bahwa semua item memiliki diskriminasi yang sama (Kelley & Schumacher, 1984).

Proses analisis dimulai dengan pengumpulan data penilaian dari asesor yang kemudian diinput ke dalam perangkat lunak R versi 4.3.3. Dalam proses iterasi, dilakukan perhitungan berulang untuk mencapai konvergensi parameter model. Pada penelitian ini, iterasi dilakukan sebanyak 41 kali hingga mencapai nilai Log-Likelihood sebesar -754.823 dengan Max-Change 0.00009. Nilai Log-Likelihood menunjukkan tingkat kecocokan antara data yang diamati dengan model yang dihipotesiskan (Haberman, 2004), sementara Max-Change yang sangat kecil mengindikasikan stabilitas model (Gutman & Palmor, 1982; Lazar et al., 2008; Raimondo et al., 2009). Setelah iterasi mencapai konvergensi, perhitungan parameter Model Rasch dilakukan untuk setiap item dalam instrumen. Parameter-parameter yang dianalisis meliputi tingkat kesulitan item (b), serta kesesuaian item dengan model yang diukur melalui parameter outfit dan infit (Christensen, 2006; Magis et al., 2014; Müller & Kreiner, 2015; Smith & Suh, 2003) Selain itu, analisis juga mencakup pemeriksaan Kurva Karakteristik Item (Item Characteristic Curve, ICC) (McCamey, 2014; Motamedi & Tkalčič, 2023) dan Fungsi Informasi Item (Item Information Function, IFF) (Penfield, 2005; Wyse & Mapuranga, 2009) untuk mengevaluasi seberapa efektif setiap item dalam mengukur performa guru dalam menstimulasi kemampuan kognitif anak.

Hasil analisis kemudian dievaluasi untuk mengidentifikasi item-item yang mungkin memerlukan revisi lebih lanjut. Parameter-parameter yang dievaluasi meliputi tingkat kesulitan item, kecocokan item dengan model, serta proporsi informasi yang diberikan oleh setiap item. Item-item yang menunjukkan penyimpangan signifikan dari ekspektasi model diidentifikasi untuk revisi atau penghapusan guna meningkatkan kualitas keseluruhan instrumen. Analisis Model Rasch 1 PL ini bertujuan untuk memastikan bahwa instrumen penilaian visitasi akreditasi PAUD valid dan reliabel dalam mengukur performa guru dalam menstimulasi kemampuan kognitif anak usia dini. Hasil penelitian ini diharapkan dapat

DOI: 10.31004/obsesi.v84.5953

memberikan rekomendasi yang berguna untuk perbaikan instrumen dan pelaksanaan penilaian akreditasi yang lebih akurat dan objektif.

Hasil dan Pembahasan

Penelitian ini bertujuan untuk menilai kualitas instrumen visitasi akreditasi PAUD dengan menggunakan Model Rasch, khususnya pada komponen stimulasi aspek kognitif anak. Hasil analisis menunjukkan bahwa sebagian besar item dalam instrumen memiliki tingkat kesesuaian yang baik dengan model, serta menunjukkan variasi tingkat kesulitan yang sesuai dengan kemampuan responden. Informasi item dan proporsi informasi yang dihasilkan mengindikasikan bahwa instrumen ini efektif dalam menangkap variasi kemampuan kognitif anak usia dini dalam rentang kemampuan yang luas. Temuan ini memberikan bukti kuat tentang validitas dan reliabilitas instrumen tersebut, sekaligus mengidentifikasi beberapa item yang memerlukan revisi lebih lanjut untuk meningkatkan kualitas keseluruhan instrumen. Diskusi ini akan menguraikan temuan utama, implikasinya dalam konteks visitasi akreditasi PAUD, saran untuk penelitian lanjutan, dan rekomendasi untuk perbaikan instrumen.

Dalam pengujian Model Rasch, iterasi merupakan proses perhitungan berulang yang dilakukan untuk mencapai konvergensi parameter model (Cohen, 1979; Zubik-Kowal, 2019). Pada penelitian ini, iterasi dilakukan sebanyak 41 kali hingga mencapai nilai Log-Likelihood sebesar -754.823 dengan Max-Change 0.00009. Nilai Log-Likelihood menunjukkan tingkat kecocokan antara data yang diamati dengan model yang dihipotesiskan; semakin tinggi nilainya (semakin mendekati nol), semakin baik kecocokannya(Jennings, 1986). Sementara itu, Max-Change menggambarkan perubahan terbesar dalam estimasi parameter antara iterasi yang berurutan. Nilai Max-Change yang sangat kecil, yakni 0.00009, menunjukkan bahwa model telah mencapai konvergensi yang stabil, dimana perubahan parameter dari satu iterasi ke iterasi berikutnya sangat minimal. Hal ini mengindikasikan bahwa proses iterasi telah cukup untuk mendapatkan estimasi parameter yang akurat dan bahwa model yang digunakan mampu merepresentasikan data dengan baik. Dengan demikian, iterasi ke-41 ini memastikan bahwa hasil analisis yang diperoleh dari Model Rasch dapat diandalkan dan valid untuk menilai kualitas instrumen visitasi akreditasi PAUD.

Setelah melakukan iterasi, perhitungan parameter Model Rasch dilakukan. Tabel 1 disajikan output R untuk pengujian tersebut.

Tabel 1. Parameter Model Rasch

	a	b	g	u
1	1	-0.3499277	0	1
2	1	-0.5680325	0	1
3	1	-1.2245005	0	1
4	1	-0.7582687	0	1
5	1	-4.2781465	0	1
6	1	-3.8609245	0	1
7	1	-0.8776684	0	1
8	1	-1.8228698	0	1
9	1	0.5697160	0	1
10	1	-3.3265741	0	1
11	1	-3.1324630	0	1
12	1	-0.4216825	0	1
13	1	0.1753190	0	1

Output hasil pengujian Model Rasch yang ditampilkan pada Tabel 1 mencakup beberapa parameter penting, yaitu a, b, g, dan u, untuk setiap item dalam instrumen yang diuji. Parameter "a" menunjukkan bahwa semua item memiliki nilai yang sama, yaitu 1, yang dalam konteks ini dapat diartikan sebagai item yang diukur pada skala yang sama atau dengan tingkat diskriminasi yang sama. Parameter "b" merepresentasikan tingkat kesulitan

DOI: 10.31004/obsesi.v84.5953

masing-masing item. Nilai "b" yang negatif menunjukkan bahwa item tersebut relatif lebih mudah dijawab dengan benar oleh responden, sementara nilai "b" yang positif menunjukkan tingkat kesulitan yang lebih tinggi (Bond et al., 2020). Misalnya, item ke-5 dengan nilai "b" sebesar -4.2781465 adalah item yang sangat mudah, sedangkan item ke-9 dengan nilai "b" sebesar 0.5697160 adalah item yang lebih sulit dibandingkan dengan yang lain. Variasi nilai "b" ini penting untuk memastikan bahwa instrumen memiliki rentang kesulitan yang luas, sehingga dapat mengukur kemampuan atau karakteristik responden secara lebih komprehensif.

Parameter "g" dan "u" dalam output tersebut masing-masing bernilai 0 dan 1 untuk semua item. Parameter "g" yang bernilai 0 dapat diinterpretasikan sebagai indikator bahwa tidak ada guessing parameter yang dipertimbangkan dalam model ini, atau bahwa tidak ada item yang mengandung kemungkinan tebak-menebak oleh responden. Sementara itu, parameter "u" yang selalu bernilai 1 menunjukkan bahwa semua item telah memenuhi criteria unidimensionality, yang berarti setiap item hanya mengukur satu kemampuan atau karakteristik yang sama. Keseragaman nilai "u" ini mengindikasikan bahwa instrumen memiliki konsistensi dalam mengukur atribut yang dimaksud. Dengan demikian, analisis Model Rasch ini memberikan gambaran bahwa instrumen penilaian akreditasi PAUD yang diuji memiliki tingkat kesulitan item yang bervariasi dan telah memenuhi asumsi dasar model Rasch terkait unidimensionality dan tidak adanya pengaruh tebak-menebak dalam respon, sehingga dapat dianggap valid dan reliabel untuk digunakan dalam penilaian.

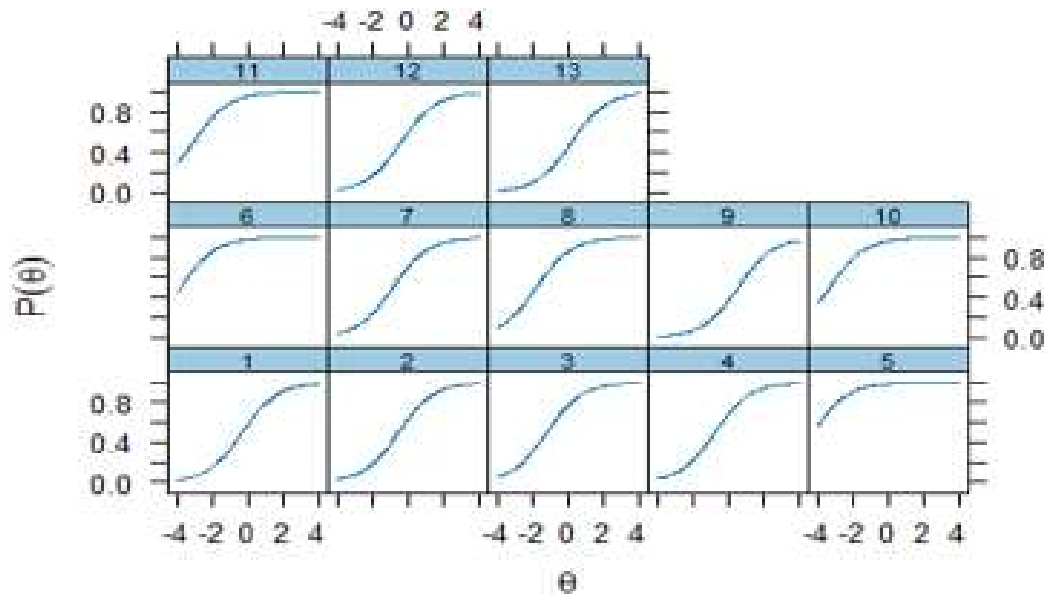
Tabel 2. Item Fit

item	outfit	z.outfit	infit	z.infit
1	0.995	-0.096	1.003	0.087
2	0.930	-1.122	0.943	-0.982
3	0.804	-1.638	0.889	-0.975
4	0.888	-1.449	0.915	-1.183
5	0.940	0.180	1.062	0.324
6	0.623	-0.490	1.043	0.267
7	0.922	-0.858	0.944	-0.672
8	0.853	-0.723	0.970	-0.113
9	0.932	-1.076	0.946	-0.969
10	0.788	-0.326	1.065	0.292
11	0.971	0.072	1.077	0.323
12	0.880	-2.409	0.896	-2.246
13	0.885	-3.149	0.893	-3.286

Tabel 2 menampilkan hasil analisis item fit menggunakan Model Rasch, yang meliputi parameter outfit dan infit serta z-scores terkait untuk setiap item dalam instrumen. Parameter outfit dan infit merupakan indikator yang digunakan untuk mengukur kesesuaian setiap item terhadap model Rasch (Meyer, 2014). Outfit merupakan ukuran sensitivitas terhadap outliers atau respon yang tidak sesuai, sementara infit lebih fokus pada kesesuaian dalam merespon item dengan tingkat kesulitan yang sesuai dengan kemampuan responden. Nilai outfit dan infit yang mendekati 1 menunjukkan kesesuaian yang baik. Dalam output ini, sebagian besar item memiliki nilai outfit dan infit yang berada dalam rentang yang dapat diterima, yaitu sekitar 1. Misalnya, item 1 memiliki nilai outfit 0.995 dan infit 1.003, yang menunjukkan kesesuaian yang baik. Namun, ada beberapa item seperti item 6 yang memiliki nilai outfit 0.623 dan item 3 dengan nilai infit 0.889 yang mengindikasikan sedikit penyimpangan, meskipun masih dalam batas toleransi.

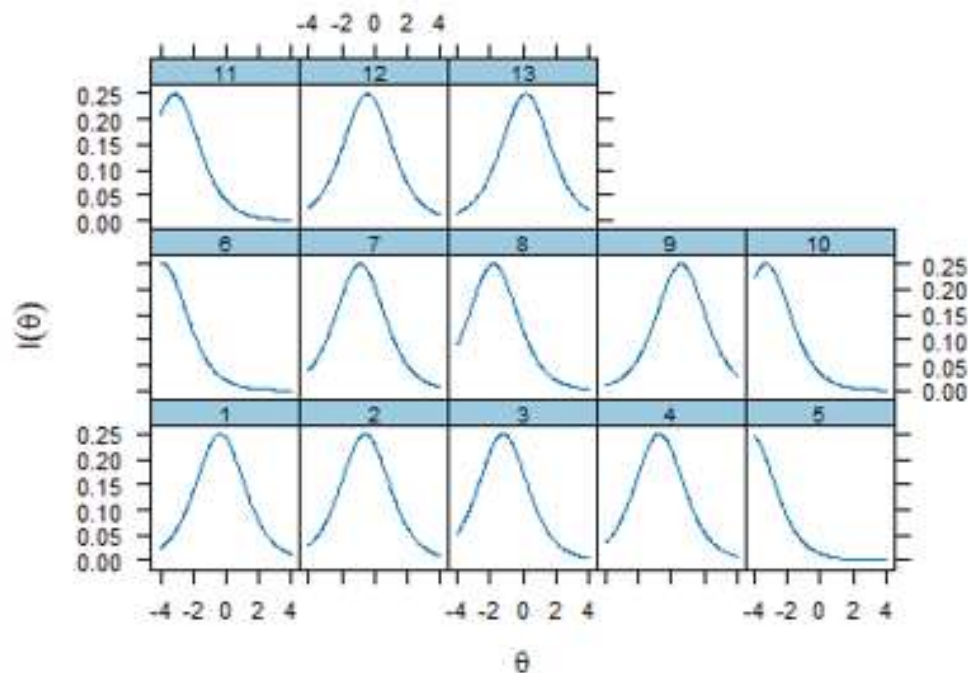
Selain itu, z-scores untuk outfit dan infit memberikan informasi tambahan tentang sejauh mana nilai-nilai tersebut menyimpang dari ekspektasi model. Z-scores yang mendekati 0 menunjukkan bahwa penyimpangan tersebut tidak signifikan secara statistik. Dalam output ini, beberapa item memiliki z-scores yang cukup besar, misalnya item 13 dengan z.outfit -3.149 dan z.infit -3.286, yang menunjukkan bahwa item ini secara signifikan menyimpang dari model yang diharapkan. Penyimpangan ini mungkin menunjukkan adanya masalah dengan

DOI: 10.31004/obsesi.v84.5953

item tersebut, seperti bias atau ketidaksesuaian dengan dimensi yang diukur. Secara keseluruhan, hasil ini mengindikasikan bahwa meskipun sebagian besar item dalam instrumen penilaian akreditasi PAUD sesuai dengan model Rasch, ada beberapa item yang memerlukan evaluasi lebih lanjut untuk memastikan validitas dan reliabilitas instrumen secara keseluruhan. Evaluasi lanjutan bisa melibatkan revisi atau penghapusan item yang menunjukkan penyimpangan signifikan untuk meningkatkan kualitas instrumen.

Gambar 1. Item Characteristic Curve (ICC)

Plot unidimensi yang ditampilkan adalah Kurva Karakteristik Item (Item Characteristic Curve, ICC) untuk setiap item dalam instrumen yang diukur menggunakan Model Rasch. Sumbu horizontal (θ) merepresentasikan kemampuan latent atau trait yang diukur, yang berkisar dari -4 hingga 4. Sumbu vertikal (P(θ)) menunjukkan probabilitas responden dengan kemampuan tertentu untuk menjawab item dengan benar. Setiap kurva menunjukkan hubungan antara kemampuan responden dan probabilitas memberikan jawaban yang benar pada item tersebut (Meyer, 2014).

Dari plot ini, dapat dilihat bahwa sebagian besar kurva menunjukkan peningkatan probabilitas jawaban benar seiring dengan meningkatnya nilai θ, yang menunjukkan bahwa instrumen secara keseluruhan berfungsi sesuai dengan prinsip model Rasch: semakin tinggi kemampuan responden, semakin besar probabilitas mereka menjawab benar. Item dengan kurva yang lebih curam menunjukkan bahwa item tersebut lebih baik dalam membedakan antara responden dengan kemampuan berbeda. Misalnya, item 11, 12, dan 13 menunjukkan kurva yang lebih tajam, yang mengindikasikan bahwa item tersebut sangat sensitif terhadap perbedaan kemampuan responden. Sebaliknya, item dengan kurva yang lebih landai, seperti item 5, menunjukkan bahwa item tersebut kurang diskriminatif. Plot ini juga menunjukkan bahwa semua item memiliki tingkat kesulitan yang berbeda-beda, yang penting untuk memastikan bahwa instrumen mampu mengukur rentang kemampuan yang luas. Secara keseluruhan, plot ICC ini memberikan visualisasi yang kuat tentang bagaimana setiap item berfungsi dalam mengukur kemampuan latent yang dimaksud dan menunjukkan bahwa instrumen memiliki variasi kesulitan yang diperlukan untuk penilaian yang komprehensif dan akurat.

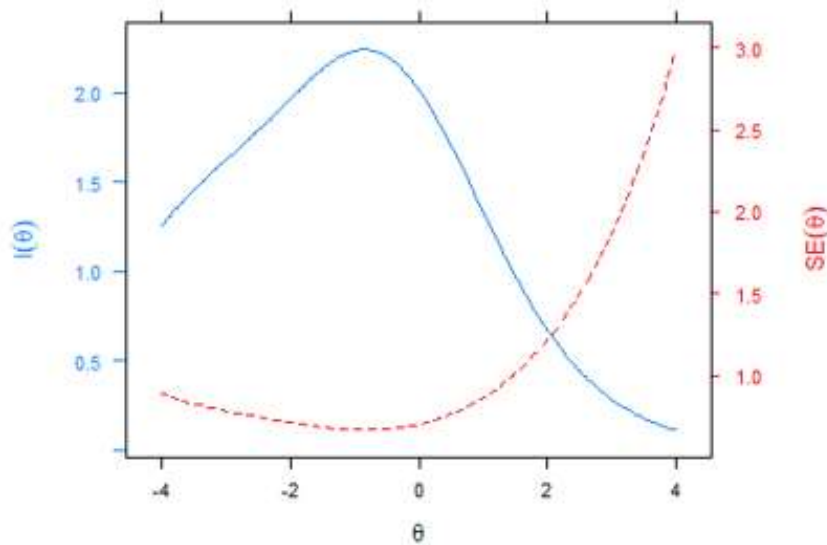
DOI: 10.31004/obsesi.v84.5953

Gambar 2. Item Information Function (IFF)

Gambar 2 menunjukkan informasi item yang dihasilkan dari Model Rasch untuk setiap item dalam instrumen. Kolom "Item" menunjukkan nomor item, "Info(-4,4)" menunjukkan informasi yang diperoleh dari item tersebut dalam rentang kemampuan θ dari -4 hingga 4, "TotalInfo" selalu bernilai 1 yang mengindikasikan bahwa seluruh informasi dari item tersebut diperhitungkan, dan "Proportion" menunjukkan proporsi informasi dari item tersebut terhadap total informasi yang ada dalam instrumen.

Informasi item adalah ukuran seberapa banyak informasi yang diberikan item tersebut mengenai kemampuan latent yang diukur. Nilai informasi yang lebih tinggi menunjukkan bahwa item tersebut lebih efektif dalam mengukur kemampuan responden dalam rentang kemampuan yang ditentukan. Dari output ini, terlihat bahwa item-item seperti item 1, 2, 4, 7, 9, 12, dan 13 memiliki nilai informasi yang tinggi, mendekati 1, yang berarti item tersebut sangat informatif dalam mengukur kemampuan responden dalam rentang θ dari

-4 hingga 4. Sebaliknya, item 5 dan item 6 memiliki nilai informasi yang lebih rendah, masing-masing 0.4306543 dan 0.5343276, yang menunjukkan bahwa kedua item ini kurang informatif dibandingkan dengan item lainnya dalam instrumen.

Proporsi informasi menunjukkan kontribusi relatif dari setiap item terhadap keseluruhan informasi dalam instrumen. Item-item dengan nilai proporsi yang lebih tinggi seperti item 1, 12, dan 13 memberikan kontribusi yang signifikan terhadap total informasi yang diperoleh dari instrumen tersebut. Sebaliknya, item 5 memiliki kontribusi yang paling kecil, yang dapat menunjukkan bahwa item ini mungkin perlu direvisi atau dievaluasi lebih lanjut untuk meningkatkan efektivitasnya dalam mengukur kemampuan responden. Secara keseluruhan, output ini memberikan gambaran yang jelas mengenai item-item mana yang paling efektif dalam memberikan informasi mengenai kemampuan responden, serta mengidentifikasi item-item yang mungkin memerlukan perbaikan

DOI: 10.31004/obsesi.v84.5953

Gambar 3. Info Estimasi Standar Error (SE)

Grafik pada Gambar 3 merupakan hasil analisis model Rasch yang mengkaji efektivitas instrumen visitasi PAUD pada aspek kognitif. Kurva biru yang mewakili fungsi informasi item (I(θ)) menunjukkan bahwa instrumen ini memiliki tingkat informasi tertinggi pada peserta didik dengan kemampuan kognitif rata-rata (θ ≈ 0). Hal ini mengindikasikan bahwa instrumen ini paling efektif dalam membedakan kemampuan kognitif peserta didik yang berada di sekitar nilai rata-rata tersebut. Informasi yang diperoleh dari instrumen ini menurun seiring dengan menjauhnya nilai θ dari rata-rata, yang menunjukkan bahwa instrumen ini kurang sensitif dalam mengukur kemampuan peserta didik yang sangat rendah atau sangat tinggi.

Kurva merah putus-putus yang menggambarkan kesalahan standar estimasi (SE(θ)) memperlihatkan bahwa estimasi kemampuan paling akurat terjadi pada guru dengan performa rata-rata, di mana kesalahan standar berada pada titik terendah. Kesalahan standar meningkat secara signifikan untuk guru dengan performa yang berada jauh di bawah atau di atas rata-rata, yang menunjukkan ketidakpastian lebih tinggi dalam estimasi kemampuan mereka. Oleh karena itu, instrumen ini lebih dapat diandalkan untuk mengukur kemampuan kognitif anak-anak yang berada di sekitar rata-rata, sedangkan untuk anak-anak dengan kemampuan yang sangat tinggi atau sangat rendah, hasil pengukuran memerlukan interpretasi yang lebih hati-hati karena tingkat ketidakpastian yang lebih besar.

Tabel 3. Test Information Function (TIF)

LowerBound	UpperBound	Info	TotalInfo	Proportion	nitems
-4	4	10.86855	13	0.8360421	13

Output pada Tabel 3 ini menyajikan informasi mengenai rentang kemampuan (θ) yang diukur oleh instrumen, serta total informasi yang diperoleh dari keseluruhan item dalam instrumen tersebut. "LowerBound" dan "UpperBound" menunjukkan batas bawah dan batas atas dari rentang kemampuan yang diukur, yaitu dari -4 hingga 4. Kolom "Info" menunjukkan jumlah total informasi yang dihasilkan oleh instrumen dalam rentang kemampuan ini, yang bernilai 10.86855.

Kolom "TotalInfo" bernilai 13, yang menunjukkan jumlah total item dalam instrumen. "Proportion" menunjukkan proporsi informasi yang diperoleh dari instrumen, yang dalam hal ini adalah 0.8360421 atau sekitar 83.6%. Ini berarti bahwa 83.6% dari total informasi yang

DOI: 10.31004/obsesi.v84.5953

mungkin diperoleh dalam rentang kemampuan ini berhasil dijaring oleh instrumen. Terakhir, kolom "nitems" menegaskan kembali jumlah item dalam instrumen, yaitu 13 item.

Interpretasi dari data ini mengindikasikan bahwa instrumen tersebut sangat informatif dalam mengukur kemampuan responden dalam rentang θ dari -4 hingga 4. Dengan proporsi informasi sebesar 83.6%, instrumen ini dapat dianggap cukup efektif dan efisien dalam memberikan informasi yang diperlukan untuk penilaian akreditasi PAUD. Informasi ini menunjukkan bahwa instrumen tersebut mampu menangkap variasi kemampuan responden dengan baik, meskipun ada ruang untuk peningkatan guna mencapai 100% proporsi informasi. Ini dapat dilakukan melalui evaluasi dan revisi item-item yang kurang informatif untuk meningkatkan akurasi dan efektivitas pengukuran.

Keseluruhan hasil analisis menunjukkan bahwa sebagian besar item dalam instrumen memiliki tingkat kesesuaian yang baik dengan model, serta menunjukkan variasi tingkat kesulitan yang sesuai dengan kemampuan responden. Informasi item dan proporsi informasi yang dihasilkan mengindikasikan bahwa instrumen ini efektif dalam menangkap variasi kemampuan guru dalam menstimulasi aspek kognitif anak usia dini dalam rentang kemampuan yang luas. Proses iterasi, yang mencapai konvergensi setelah 41 kali iterasi dengan nilai Log-Likelihood sebesar -754.823 dan Max-Change 0.00009, memastikan stabilitas dan akurasi estimasi parameter, sehingga dapat memastikan bahwa model yang digunakan mampu merepresentasikan data dengan baik. Parameter item mengungkapkan variasi tingkat kesulitan, dengan item yang lebih mudah (misalnya item 5 dengan tingkat kesulitan -4.2781465) dan yang lebih sulit (misalnya item 9 dengan tingkat kesulitan 0.5697160). Analisis kecocokan item menunjukkan bahwa meskipun sebagian besar item menunjukkan statistik kecocokan yang dapat diterima, beberapa item memerlukan evaluasi lebih lanjut untuk kemungkinan revisi.

Temuan ini mendukung validitas dan reliabilitas instrumen untuk menilai performa guru dalam menstimulasi kognitif anak di lingkungan PAUD. Variasi tingkat kesulitan item penting untuk memastikan bahwa instrumen dapat menangkap berbagai kemampuan guru, memastikan penilaian yang akurat untuk guru dengan performa yang rendah maupun tinggi(Purniati et al., 2020). Identifikasi item dengan penyimpangan signifikan, seperti item 13 yang menunjukkan skor-z yang mencolok untuk outfit dan infit, menunjukkan area di mana instrumen dapat diperbaiki untuk meningkatkan kualitas pengukuran keseluruhan. Implikasinya, administrator pendidikan dan pembuat kebijakan dapat menggunakan instrumen ini dengan keyakinan bahwa instrumen memberikan evaluasi komprehensif terhadap praktik stimulasi kognitif.

Penelitian selanjutnya perlu difokuskan pada revisi item yang menunjukkan penyimpangan signifikan dari ekspektasi model. Hal ini dapat melibatkan studi kualitatif untuk memahami mengapa item tertentu tidak sesuai, serta mengeksplorasi pengungkapan atau konsep alternatif yang mungkin lebih baik dalam menangkap keterampilan kognitif yang dimaksud. Selain itu, memperluas ukuran dan keragaman sampel dapat memberikan validasi yang lebih kuat terhadap instrumen di berbagai konteks dan populasi. Studi longitudinal juga dapat meneliti efektivitas instrumen dalam melacak perkembangan kognitif dari waktu ke waktu, memberikan wawasan tentang dampak jangka panjang dari intervensi pendidikan.

Rekomendasi yang diusulkan mencakup revisi item yang menunjukkan masalah kecocokan signifikan seperti item 13, serta memberikan pelatihan menyeluruh kepada pendidik dan penilai tentang cara mengadministrasikan instrumen secara akurat dan menginterpretasikan hasilnya dengan efektif. Uji coba item yang direvisi perlu dilakukan sebelum implementasi untuk memastikan peningkatan kesesuaian dan fungsionalitas. Selain itu, perlu dipertimbangkan untuk memperluas instrumen agar mencakup keterampilan kognitif tambahan atau mengintegrasikannya dengan domain perkembangan lainnya seperti perkembangan sosial-emosional atau fisik untuk penilaian yang lebih holistik.

DOI: 10.31004/obsesi.v84.5953

Hasil penelitian ini menunjukkan bahwa penerapan Model Rasch dalam menilai instrumen visitasi akreditasi PAUD memberikan wawasan berharga mengenai kualitas dan efektivitas instrument tersebut. Meskipun instrumen menunjukkan validitas dan reliabilitas yang kuat secara keseluruhan, revisi yang ditargetkan dan penelitian lebih lanjut akan membantu meningkatkan utilitas dan akurasinya. Upaya ini pada akhirnya akan berkontribusi pada evaluasi yang lebih kuat dan komprehensif terhadap program pendidikan anak usia dini, mendukung tujuan untuk membina warga negara yang dinamis dan berbakat di masa depan.

Simpulan

Hasil analisis menunjukkan bahwa sebagian besar item dalam instrumen visitasi akreditasi PAUD memiliki tingkat kesesuaian yang baik dengan model dan variasi tingkat kesulitan yang sesuai dengan kemampuan responden. Informasi yang dihasilkan oleh instrumen menunjukkan efektivitasnya dalam menangkap variasi performa guru dalam menstimulasi kemampuan kognitif anak usia dini, dengan bukti kuat tentang validitas dan reliabilitas instrumen tersebut. Meskipun demikian, beberapa item teridentifikasi memerlukan revisi lebih lanjut untuk meningkatkan kualitas keseluruhan instrumen. Item 3, 6, dan 13 menunjukkan ketidaksesuaian dengan model atau sensitivitas terhadap outliers, yang menandakan perlunya evaluasi dan perbaikan. Revisi yang disarankan mencakup evaluasi desain item, penyederhanaan bahasa, pelatihan asesor, reformulasi pertanyaan, peningkatan instruksi, serta validasi ulang dengan ahli dan pengujian ulang. Dengan melakukan langkah-langkah perbaikan ini, diharapkan instrumen penilaian akreditasi PAUD dapat memberikan hasil yang lebih objektif, valid, dan reliabel, serta mendukung peningkatan kualitas pendidikan anak usia dini di Indonesia.

Daftar Pustaka

A Aziz, N. F., Ahmad, H., & Mat Nashir, I. (2019). Validation of technical and vocational teachers' competency evaluation instrument using the Rasch model. *Jurnal Pendidikan Sains Dan Matematik Malaysia*, *9*(1), 18–25. https://doi.org/10.37134/jpsmm.vol9.1.3.2019

Applications of Rasch Measurement in Learning Environments Research. (2011). In R. F. Cavanagh & R. F. Waugh (Eds.), *Applications of Rasch Measurement in Learning Environments Research*. SensePublishers. https://doi.org/10.1007/978-94-6091-493-5

Bond, T., Yan, Z., & Heene, M. (2020). Applying the Rasch Model. In *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Routledge. https://doi.org/10.4324/9780429030499

Boone, W., & Rogan, J. (2005). Rigour in quantitative analysis: The promise of rasch analysis techniques. *African Journal of Research in Mathematics, Science and Technology Education*, *9*(1), 25–38. https://doi.org/10.1080/10288457.2005.10740574

Bui, T. L. T., Kazarenkov, V. I., & de Tran, V. (2020). Application of rasch model to develop a questionnaire for evaluating the quality of teaching for students' creativity development. *International Journal of Learning, Teaching and Educational Research*, *19*(8), 278–296. https://doi.org/10.26803/ijlter.19.8.15

Christensen, K. B. (2006). Fitting polytomous Rasch models in SAS. *Journal of Applied Measurement*, *7*(4), 407–417.

Clark, K. E., & Kriedt, P. H. (1948). An application of Guttman's new scaling techniques to an attitude questionnaire. *Educational and Psychological Measurement*, *8*(2), 215–223. https://doi.org/10.1177/001316444800800206

Cohen, L. (1979). Approximate expressions for parameter estimates in the Rasch model. *British Journal of Mathematical and Statistical Psychology*, *32*(1), 113–120. https://doi.org/10.1111/j.2044-8317.1979.tb00756.x

De Morton, N., & Keating, J. (2008). Health instruments itching for a Rasch. *International Journal*

of Therapy and Rehabilitation, *15*(2), 56. https://doi.org/10.12968/ijtr.2008.15.2.28186

Frey, B. B. (2018). Guttman Scaling. In *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications, Inc. https://doi.org/10.4135/9781506326139.n298

Gutman, S., & Palmor, Z. (1982). Properties of Min-Max Controllers in Uncertain Dynamical Systems. *SIAM Journal on Control and Optimization*, *20*(6), 850–861. https://doi.org/10.1137/0320060

Haberman, S. J. (2004). Joint and Conditional Maximum Likelihood Estimation for the Rasch Model for Binary Responses. *ETS Research Report Series*, *2004*(1), i–63. https://doi.org/10.1002/j.2333-8504.2004.tb01947.x

Jennings, D. E. (1986). Judging inference adequacy in logistic regression. *Journal of the American Statistical Association*, *81*(394), 471–476. https://doi.org/10.1080/01621459.1986.10478292

Karlin, O., & Karlin, S. (2018). Making Better Tests with the Rasch Measurement Model. *InSight: A Journal of Scholarly Teaching*, *13*, 76–100. https://doi.org/10.46504/14201805ka

Kelley, P. R., & Schumacher, C. F. (1984). The rasch model: Its Use by the National Board of Medical Examiners. *Evaluation & the Health Professions*, *7*(4), 443–454. https://doi.org/10.1177/016327878400700405

Kemendikbud. (2022). *Kajian Akademik Evaluasi Sistem Pendidikan*. 1–49.

Krishnan, S., & Idris, N. (2018). Using Partial Credit Model to Improve the Quality of an Instrument. *International Journal of Evaluation and Research in Education (IJERE)*, *7*(4), 313. https://doi.org/10.11591/ijere.v7i4.15146

Lash, T. L., Ahern, T. P., Collin, L. J., Fox, M. P., & MacLehose, R. F. (2021). Bias Analysis Gone Bad. *American Journal of Epidemiology*, *190*(8), 1604–1612. https://doi.org/10.1093/aje/kwab072

Lazar, M., Muñoz de la Peña, D., Heemels, W. P. M. H., & Alamo, T. (2008). On input-to-state stability of min-max nonlinear model predictive control. *Systems and Control Letters*, *57*(1), 39–48. https://doi.org/10.1016/j.sysconle.2007.06.013

Magis, D., Beland, S., & Raiche, G. (2014). Snijders's correction of Infit and Outfit indexes with estimated ability level: an analysis with the Rasch model. *Journal of Applied Measurement*, *15*(1), 82–93.

McCamey, R. (2014). A Primer on the One-Parameter Rasch Model. *American Journal of Economics and Business Administration*, *6*(4), 159–163. https://doi.org/10.3844/ajebasp.2014.159.163

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for Response Bias as a Source of Error Variance in Applied Assessment. *Psychological Bulletin*, *136*(3), 450–470. https://doi.org/10.1037/a0019216

Meyer, J. P. (2014). Rasch Measurement. In *Applied Measurement with jMetrik* (pp. 102–117). Routledge. https://doi.org/10.4324/9780203115190-14

Mohamad, M. M., Sulaiman, N. L., Sern, L. C., & Salleh, K. M. (2015). Measuring the Validity and Reliability of Research Instruments. *Procedia - Social and Behavioral Sciences*, *204*, 164–171. https://doi.org/10.1016/j.sbspro.2015.08.129

Motamedi, E., & Tkalčič, M. (2023). User-centric item characteristics for personalized multimedia systems: A systematic review. *Intelligenza Artificiale*, *17*(2), 207–228. https://doi.org/10.3233/IA-230039

Müller, M., & Kreiner, S. (2015). *Item Fit Statistics in Common Software for Rasch Analysis*.

Penfield, R. D. (2005). Unique properties of Rasch model item information functions. *Journal of Applied Measurement*, *6*(4), 355–365.

Pillet, J. C., Vitari, C., Pigni, F., & Carillo, K. (2018). Detecting biased items when developing a scale: A quantitative approach. *Americas Conference on Information Systems 2018: Digital Disruption, AMCIS 2018*.

Purniati, T., Turmudi, Evayanti, M., & Suhaedi, D. (2020). Analysis of Instruments and Students' Abilities in Material Number Patterns Using the Rasch Model. *Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019*. https://doi.org/10.4108/eai.12-10-2019.2296432

Raimondo, D. M., Limon, D., Lazar, M., Magni, L., & Camacho, E. F. (2009). Min-max model predictive control of nonlinear systems: A unifying overview on stability. *European Journal of Control*, *15*(1), 5–21. https://doi.org/10.3166/ejc.15.5-21

Roßbach, H.-G. (1990). *Assessing the Quality of Kindergarten Environments with the Early Childhood Environment Rating Scale* (pp. 77–90). https://doi.org/10.1007/978-3-642-84256-6_7

Sirodj, D. A. N., Permana, R. H., & Suhana, S. (2020). Application of the Rasch Model in Analysis of Exam Questions at the Faculty of Psychology of Universitas Islam Bandung. *Proceedings of the 2nd Social and Humaniora Research Symposium (SoRes 2019)*. https://doi.org/10.2991/assehr.k.200225.059

Smith, R. M., & Suh, K. K. (2003). Rasch Fit Statistics as a Test of the Invariance of Item Parameter Estimates. *Journal of Applied Measurement*, *4*(2), 153–163.

Umobong, M. E. (2017). The One-Parameter Logistic Model (1PLM) And Its Application In Test Development. *Advances in Social Sciences Research Journal*, *4*(24), 126–137. http://dx.doi.org/10.14738/assrj.424.3938.

Wilkerson, J. R., & Lang, W. S. (2006). Measuring teaching ability with the Rasch model by scaling a series of product and performance tasks. *Journal of Applied Measurement*, *7*(3), 239–259.

Wyse, A. E., & Mapuranga, R. (2009). Differential Item Functioning Analysis Using Rasch Item Information Functions. *International Journal of Testing*, *9*(4), 333–357. https://doi.org/10.1080/15305050903352040

Yan, Z. (2010). Objective measurement in psychological science: an overview of Rasch model. *Advances in Psychological Science*, *18*(08), 1298–1305.

Zubik-Kowal, B. (2019). On the Convergence of Dynamic Iterations in Terms of Model Parameters. In *Integral Methods in Science and Engineering: Analytic Treatment and Numerical Approximations* (pp. 465–476). Springer International Publishing. https://doi.org/10.1007/978-3-030-16077-7_36